

Perbandingan Sistem Perhitungan Suara Tepuk Tangan dengan Metode Berbasis Frekuensi dan Metode Berbasis Amplitudo

Comparison of Applause Calculation Systems using Frequency-Based Method and Amplitude-Based Method

PUSPITA KARTIKA SARI*, KARLISA PRIANDANA, AGUS BUONO

Abstrak

Sistem penilaian berdasarkan suara tepuk tangan sering digunakan dalam acara perlombaan di Indonesia. Namun, penentuan pemenang dengan cara konvensional cenderung subjektif. Penelitian ini mengembangkan sistem penilaian otomatis berbasis komputer untuk menghitung jumlah orang bertepuk tangan dan menentukan pemenang dari perlombaan berdasarkan tepuk tangan. Penelitian ini membandingkan dua metode yang dapat diterapkan yaitu metode berbasis frekuensi dan metode berbasis amplitudo. Metode yang berbasis frekuensi mengimplementasikan *Mel Frequency Cepstral Coefficient* (MFCC) sebagai pengekstraksi ciri dan *codebook* sebagai pengenalan pola. Hasil yang diperoleh merupakan suatu model berupa kelas-kelas yang diklusterkan oleh *K-Means clustering*. Parameter penting dalam metode ini adalah jumlah koefisien *cepstral*, *overlap*, *time frame*, dan jumlah kluster. Beberapa pengujian dilakukan untuk menemukan parameter optimum dengan nilai akurasi tertinggi. Metode kedua merupakan metode berbasis amplitudo yang dilakukan dengan menghitung jumlah sampel sinyal yang memiliki nilai amplitudo di atas nilai-nilai ambang (*thresholds*) tertentu yang menghasilkan akurasi maksimum. Hasil penelitian menunjukkan bahwa akurasi sistem berbasis frekuensi untuk tepuk tangan periodik adalah 83.3% dan untuk tepuk tangan acak ialah 50% sehingga akurasi sistem untuk tepuk tangan acak berbasis *threshold* yang lebih sederhana ialah 66.7%. Dengan demikian, metode berbasis amplitudo baik digunakan.

Kata kunci: *Codebook*, *K-means*, *Mel Frequency Cepstral Coefficients* (MFCC), Pengenalan Suara, *Threshold*

Abstract

Scoring system based on sound of applause is often used in many contests in Indonesia. However, determining the winner in a conventional way tend to be subjective. This research develops automated computer-based scoring system to count the number of people who applauded and determine the winner of a competition based on applause. This research compares two methods that can be adopted: frequency-based method and amplitude-based method. Frequency-based method utilizes Mel Frequency Cepstral Coefficient (MFCC) for feature extraction and Codebook for pattern recognition. They produce a model in the form of classes that are clustered by K-Means clustering. The important parameters in this method are the number of cepstral coefficients, overlap, the time frame, and the number of clusters. Several tests are conducted to find the optimum parameters with maximum accuracy. The second method is amplitude-based method by counting the number of signal samples with amplitude above a certain threshold value that can produce the maximum accuracy. It is found that the accuracy of frequency-based system for periodic applause is 83.3%, whereas that for real applause is 50%. Meanwhile, the accuracy of a simpler threshold-based system for real applause is 66.7%. Thus, this study recommends the usage of amplitude-based method for this application.

Keywords: *Codebook*, *K-means*, *Mel Frequency Cepstral Coefficients* (MFCC), *Speaker Recognition*, *Threshold*

PENDAHULUAN

Sistem penilaian berdasarkan suara tepuk tangan sering digunakan dalam acara perlombaan di Indonesia. Pemenang ditentukan berdasarkan suara tepuk tangan terbanyak dari para penontonnya, namun pembawa acara, juri, atau siapapun yang menentukan pemenang berdasarkan tepuk tangan cenderung subjektif. Penggunaan komputer sebagai

sistem penilaian merupakan salah satu solusi yang dapat dilakukan, karena sifat komputer yang tidak subjektif dan konsisten terhadap tugas yang diperintahkan.

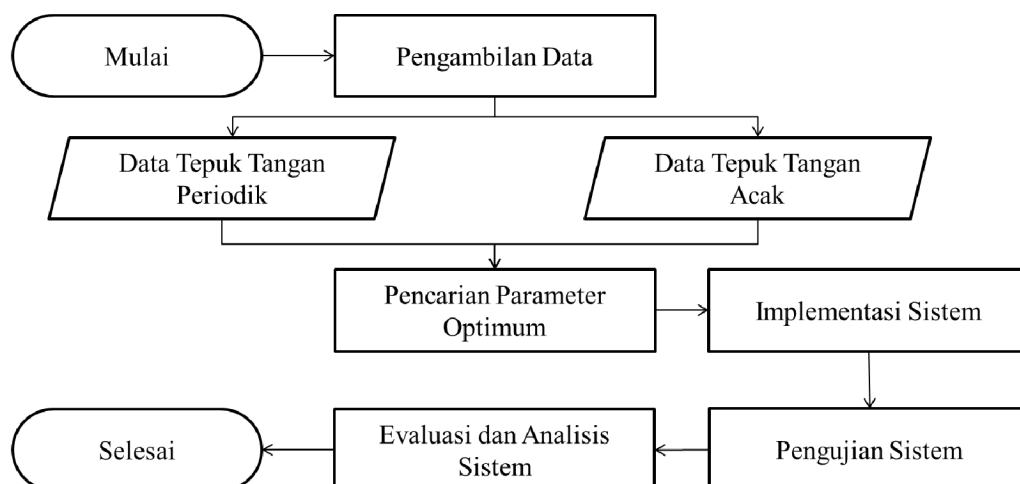
Pada proses pengenalan jumlah orang yang bertepuk tangan, salah satu fitur yang penting untuk dianalisis adalah frekuensi. Metode *Mel Frequency Cepstral Coefficients* (MFCC), *Linear Predictive Cepstral Coefficients* (LPCC), dan Wavelet sering digunakan untuk mengekstraksi fitur dalam domain frekuensi pada suara. Metode LPCC memiliki keunggulan dalam hal komputasi yang lebih sederhana, namun tingkat akurasi tidak sebaik MFCC (Li dan Chang 2003). Pernyataan ini diperkuat bahwa, dalam pengenalan suara, MFCC dapat merepresentasikan sinyal lebih baik dibandingkan dengan LPCC dan teknik lainnya (Ganchev 2005). Penelitian lain menyimpulkan bahwa penggunaan metode Wavelet sebagai ekstraksi ciri pada pengenalan pola tidak memberikan hasil yang lebih baik daripada penggunaan metode MFCC (Taufani 2011). Oleh karena itu, metode MFCC digunakan pada penelitian ini sebagai metode berbasis frekuensi untuk mengekstraksi ciri. Kelebihan utama dari MFCC ialah dapat meniru perilaku dari telinga manusia. Setelah ekstraksi ciri, langkah lain yang diperlukan ialah pengenalan pola. Pada penelitian ini, metode *codebook* akan digunakan untuk melakukan pengenalan pola.

Selain frekuensi, fitur lain yang penting untuk diamati pada pengenalan jumlah orang yang bertepuk tangan ialah amplitudo. Dalam penelitian ini metode berbasis amplitudo adalah metode *threshold*, dengan cara menghitung jumlah sampel sinyal suara yang memiliki amplitudo di atas nilai tertentu.

Sistem yang dikembangkan dalam penelitian ini ialah *prototype* dengan fokus utama pada tahapan pemodelan dan pengujian identifikasi suara tepuk tangan. *Prototype* yang dikembangkan hanya dapat melakukan identifikasi suara tepuk tangan maksimal untuk 10 orang. *Prototype* ini diharapkan dapat menjadi dasar pengembangan sistem identifikasi suara tepuk tangan.

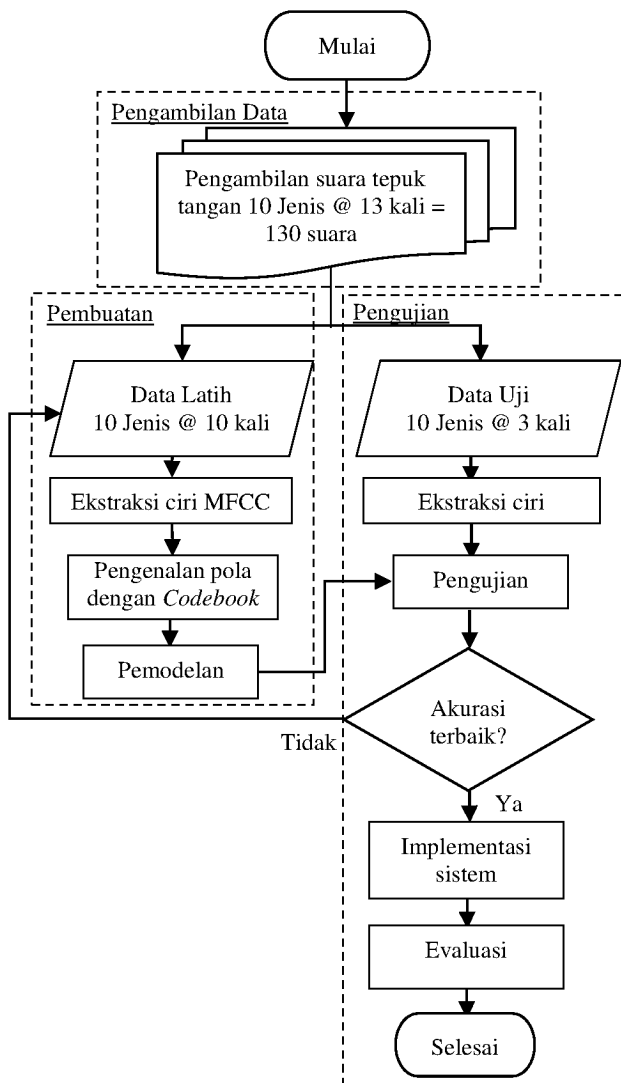
METODE

Pada penelitian ini, terdapat dua metode utama yang diimplementasikan yaitu metode berbasis frekuensi dan metode berbasis amplitudo. Alur penelitian secara umum ditunjukkan pada Gambar 1.

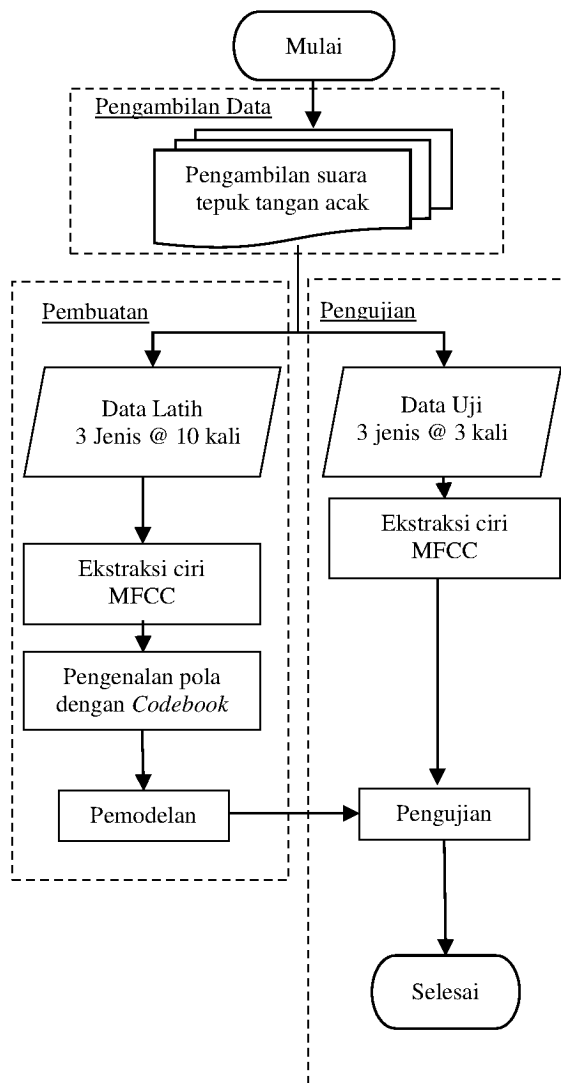


Gambar 1 Alur penelitian

Untuk sistem berbasis frekuensi yang lebih kompleks, terdapat dua subalur yang terdiri atas pemrosesan data tepuk tangan periodik (Gambar 2) dan pemrosesan data tepuk tangan acak (Gambar 3). Data tepuk tangan acak diperlukan agar sistem dapat diimplementasikan untuk mengenali tepuk tangan secara *real time*.



Gambar 2 Proses sistem penilaian otomatis dengan data tepuk tangan periodik untuk mencari parameter optimum



Gambar 3 Proses pembuatan sistem penilaian otomatis dengan data tepuk tangan acak

Data Tepuk Tangan Periodik

Data suara yang digunakan pada penelitian ini berasal dari 10 orang. Terdapat 10 jenis kelas tepuk tangan yang diambil, yaitu kelas tepuk tangan 1 dengan hanya satu orang yang bertepuk tangan, kelas tepuk tangan 2 dengan dua orang bertepuk tangan, dan seterusnya hingga kelas tepuk tangan 10 dengan semua (10) orang bertepuk tangan. Tiap kelas tepuk tangan direkam sebanyak 13 kali, masing-masing dalam waktu yang berbeda. *File* data suara disimpan dalam bentuk fail WAV dan setiap perekaman suara menggunakan *sampling rate* 11000 Hz. Dari 13 data suara dari masing-masing kelas, sebanyak 10 data dijadikan data latih dan 3 data suara lainnya dijadikan data uji.

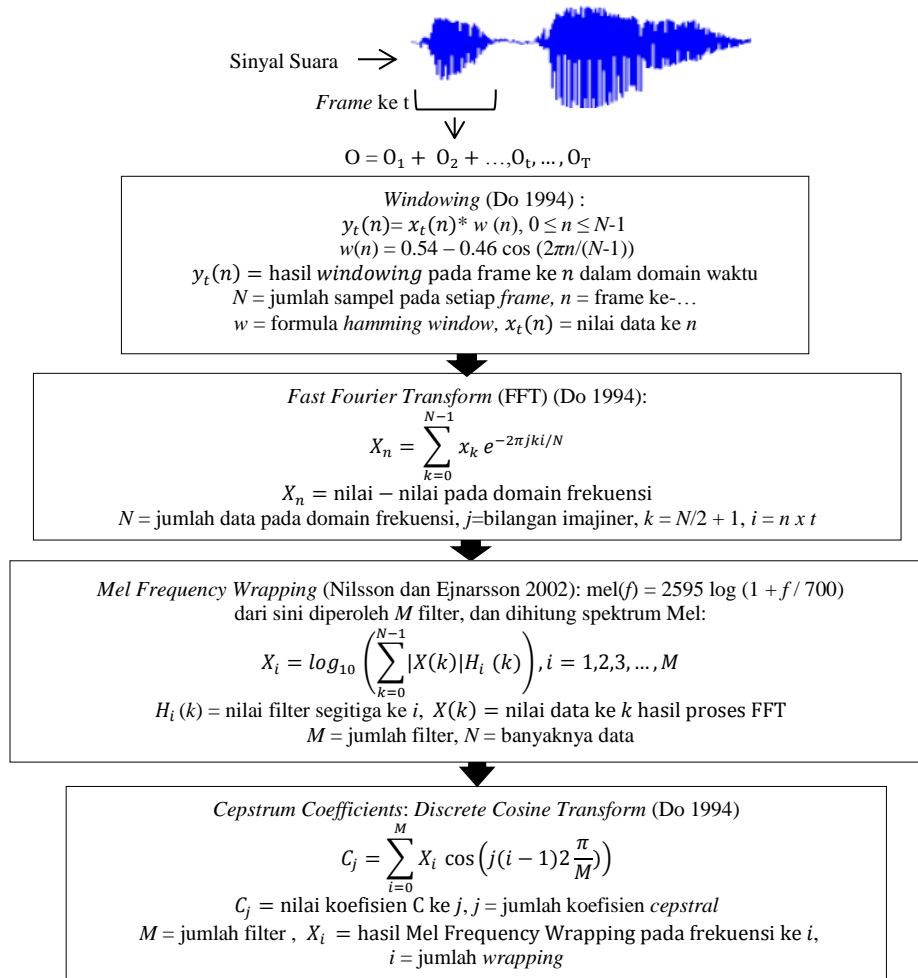
Data latih diproses dengan ekstraksi ciri MFCC, pengenalan pola *codebook*, dan pemodelan *K-means clustering* untuk menghasilkan suatu model sebagai alat penentu keputusan dari hasil pengujian. Setelah itu, data uji akan diproses dengan ekstraksi ciri MFCC kemudian dicocokkan dengan model sehingga didapatkan hasil klasifikasi tepuk tangan.

Data Tepuk Tangan Acak

Kelas yang dibuat berupa rentang, yaitu kelas 1 dengan 0-2 orang bertepuk tangan, kelas 2 dengan 3-7 orang bertepuk tangan, dan kelas 3 dengan lebih dari 7 orang bertepuk tangan. Masing-masing kelas memiliki 13 data yang dibagi menjadi 10 data latih dan 3 data uji.

Ekstraksi Ciri dengan MFCC

Gambar 4 menjelaskan mengenai tahapan ekstraksi ciri dengan teknik MFCC (Buono 2009).



Gambar 4 Diagram alur *mel frequency cepstral coefficients*

Pada proses *frame blocking*, sinyal suara disegmentasi menjadi beberapa *frame* yang saling tumpang tindih (*overlap*) (Gambar 4) agar tidak ada sedikitpun sinyal yang hilang (*deletion*). Sinyal dibaca dari *frame* ke *frame* dengan nilai *overlap* tertentu lalu dilakukan *windowing* untuk setiap *frame*. Selanjutnya, transformasi *Fourier* dilakukan untuk mengubah dimensi suara dari domain waktu ke domain frekuensi. Dari hasil transformasi *Fourier*, *spectrum mel* dihitung menggunakan sejumlah filter yang dibentuk untuk mengikuti persepsi sistem pendengaran manusia. Proses ini dikenal dengan *Mel Frequency Wrapping*. Koefisien MFCC atau koefisien *cepstrum* merupakan hasil transformasi *Cosinus* dari *spectrum mel* yang diperoleh.

Pengenalan Pola dengan *Codebook*

Pengenalan pola dengan *codebook* dilakukan untuk data latih, setelah vektor ciri diperoleh dari proses MFCC. *Codebook* dibuat untuk setiap data yang telah direkam dan terdiri atas beberapa *codeword* untuk merepresentasikan ciri suara. *Codebook* dibentuk

dengan cara membentuk klaster atau kelompok dari semua vektor ciri data latih dengan menggunakan suatu *clustering algorithm*. Algoritma *clustering* yang digunakan adalah *K-means*.

Pengujian

Pengujian dilakukan pada ciri data uji yang dihasilkan dari proses ekstraksi ciri MFCC dengan menggunakan model *codebook* dari data latih. *Output* yang dihasilkan berupa jumlah orang yang bertepuk tangan. Tingkat akurasi sistem dihitung untuk mengevaluasi hasil penelitian. Persentase tingkat akurasi dihitung dengan rumus sebagai berikut:

$$\text{hasil akurasi} = \frac{\text{jumlah suara yang benar}}{\text{jumlah suara yang diuji}} \times 100\%$$

Pengujian dilakukan berulang-ulang dengan mengubah nilai-nilai parameter (jumlah koefisien *cepstral*, *overlap*, *time frame*, dan jumlah klaster) untuk mendapatkan hasil akurasi maksimum.

HASIL DAN PEMBAHASAN

Pengembangan Sistem Berbasis Frekuensi

Terdapat lima parameter yang harus dimasukkan pada fungsi MFCC yang dibuat, yaitu vektor suara, *sampling rate*, *time frame*, *overlap*, dan jumlah *cepstral coefficient*. Vektor suara diperoleh dari data suara yang diubah menjadi vektor. Dalam hal ini, *sampling rate* yang digunakan adalah 11 000 Hz. Nilai parameter lainnya (*time frame*, *overlap*, dan jumlah *cepstral coefficient*) diubah-ubah hingga diperoleh nilai parameter-parameter yang optimum.

Pada proses pembuatan model *codebook*, data yang digunakan adalah data latih yang sudah berupa vektor ciri, hasil dari tahap MFCC. Data latih tersebut kemudian dikelompokkan dengan menggunakan *K-means clustering*. Jumlah klaster yang diujicobakan pada proses *K-means* ialah 2, 3, 4, dan 5.

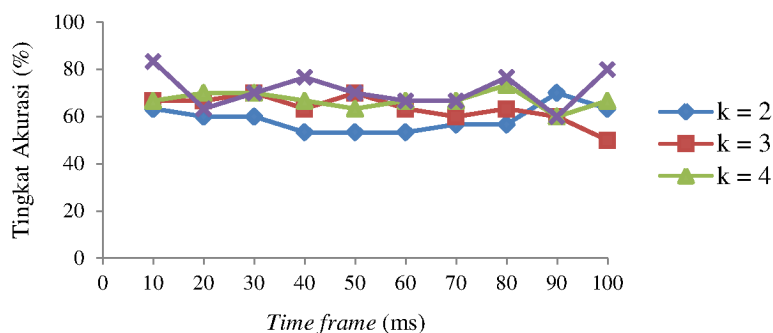
Pada penelitian ini, parameter-parameter yang diujicobakan ialah:

- 1 *time frame* sebesar {10, 20, 30, 40, 50, 60, 70, 80, 90, 100} ms,
- 2 jumlah klaster pada *K-means* sebanyak 2, 3, 4, 5,
- 3 jumlah koefisien *cepstral* sebanyak 13 dan 26, dan
- 4 *overlap* sebesar 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.6, 0.75.

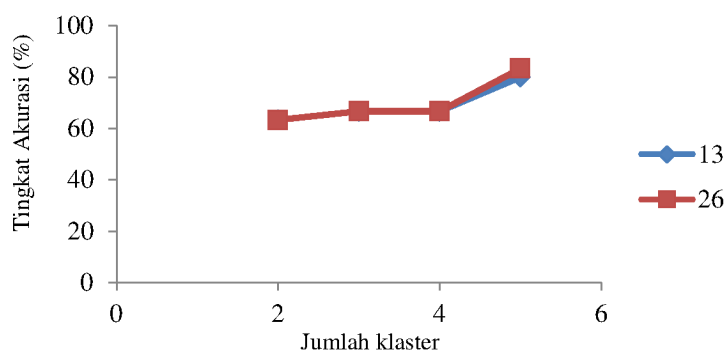
Pengaruh *time frame* terhadap tingkat akurasi (Gambar 5) tidak terlihat jelas. Sementara itu, semakin banyak jumlah klaster, maka semakin baik akurasi yang dihasilkan. Akurasi pada saat klaster sebanyak 5 dan *time frame* sebesar 10 ms menghasilkan nilai akurasi terbaik (83.3%).

Jumlah koefisien *cepstral* dan *frame* menampung vektor ciri suara. Ketika koefisien sebesar 13, akurasi maksimum yang dihasilkan yaitu 80% sedangkan jumlah koefisien *cepstral* sebesar 26 menghasilkan akurasi maksimum sebesar 83.3% (Gambar 6). Jadi, nilai koefisien tidak terlalu berpengaruh.

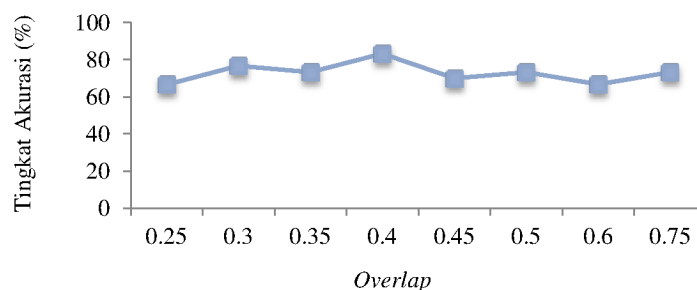
Tingkat akurasi maksimum berada pada saat *overlap* sebanyak 0.4, sedangkan, akurasi minimum yaitu 63.3% ketika *overlap* sebesar 0.25 dan 0.6 (Gambar 7). Nilai *overlap* 0.25 dianggap terlalu sedikit menangkap informasi yang hilang ketika proses *frame blocking*, sebaliknya dengan *overlap* 0.6 akan terlalu banyak mengambil informasi suara yang justru tidak diperlukan.



Gambar 5 Pengaruh nilai *time frame* dan jumlah kluster pada tingkat akurasi dengan jumlah koefisien *cepstral* 10 ms dan *overlap* 0.4



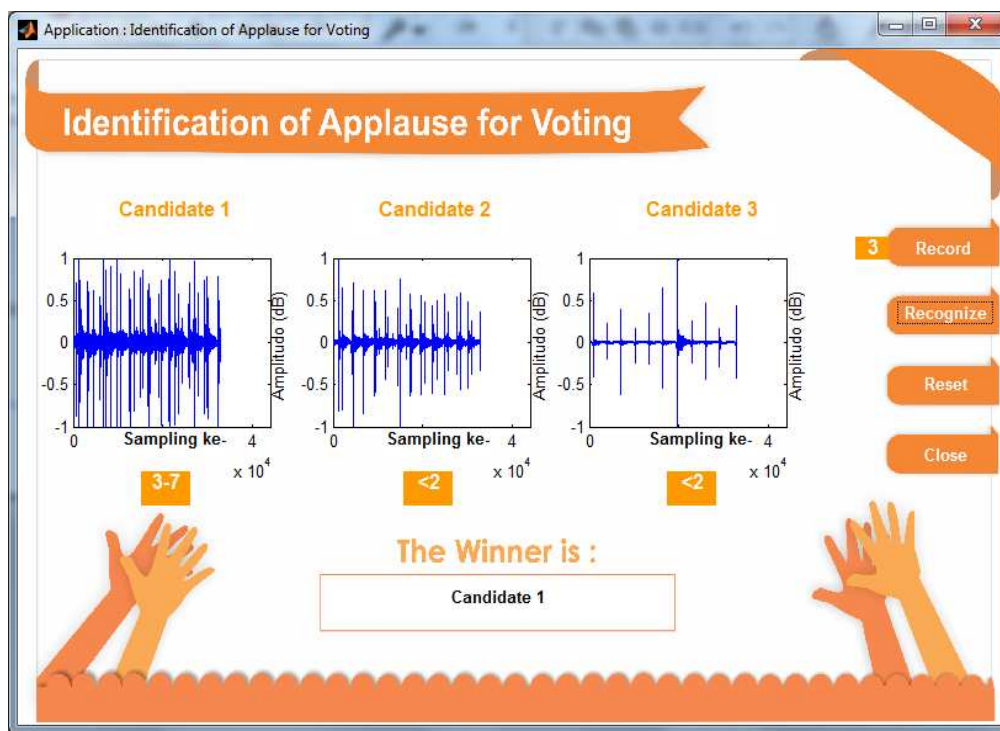
Gambar 6 Pengaruh nilai koefisien *cepstral* terhadap tingkat akurasi dengan *overlap* 0.4



Gambar 7 Pengaruh *overlap* pada tingkat akurasi dengan jumlah koefisien *cepstral* 26, jumlah kluster 5, dan *time frame* 10 ms

Sistem dibuat untuk tepuk tangan acak dengan menggunakan parameter optimum. Gambar 8 merupakan antarmuka sistem yang memiliki 3 fungsi utama, yaitu *record*, *recognize*, dan *reset*. Fungsi *record* dibuat untuk merekam suara tepuk tangan, fungsi *recognize* dibuat untuk mengidentifikasi suara tepuk tangan, dan *reset* untuk mengembalikan antarmuka seperti awal. Tiga data uji dibandingkan untuk menentukan pemenang. Identifikasi ditampilkan dalam rentang nilai 0-2 orang, 3-7 orang, dan lebih dari 7 orang.

Walaupun pencarian parameter optimum (dengan data tepuk tangan periodik) menghasilkan akurasi 83.3%, hasil pengujian sistem dengan tepuk tangan secara langsung (*real time*) hanya 50% (Tabel 1). Pengujian langsung ini dilakukan sebanyak 10 kali untuk setiap kelas.



Gambar 8 Tampilan antarmuka Sistem Identifikasi Tepuk Tangan

Tabel 1 Matriks konfusi hasil pengujian langsung data tepuk tangan acak dengan MFCC

Prediksi Aktual				
	Kelas 1	Kelas 2	Kelas 3	
Kelas 1	4	6	0	
Kelas 2	2	6	2	
Kelas 3	0	5	5	

Keterangan: Akurasi = $\frac{15}{30} \times 100\% = 50\%$

Beberapa faktor yang menyebabkan penurunan hasil akurasi pada pengujian langsung ialah adanya *noise* yang disebabkan oleh lingkungan, variasi setiap suara tepuk tangan, distorsi alat/*hardware*, dan terbatasnya jumlah data latih. Faktor lain yang menyebabkan turunnya hasil akurasi ialah pemilihan metode pengenalan suara. Metode MFCC menggunakan *power spectrum* sebagai penentu ciri. *Power spectrum* bersifat sensitif terhadap gangguan *noise*. Dari fakta ini dapat disimpulkan bahwa rendahnya akurasi disebabkan oleh kelemahan dari metode yang digunakan untuk menangkap ciri sinyal. Oleh karena itu, percobaan menggunakan metode sederhana berbasis amplitudo dengan menentukan suatu nilai *threshold* dilakukan.

Pengembangan Sistem Berbasis *Threshold*

Metode ini diimplementasikan dengan cara menghitung jumlah sampel sinyal yang memiliki nilai amplitudo lebih dari suatu *threshold* tertentu. Dalam hal ini, beberapa nilai *threshold* diujicobakan hingga diperoleh nilai *threshold* optimum yang menghasilkan akurasi maksimum. Sama seperti pada metode berbasis frekuensi dengan tepuk tangan acak, percobaan yang dilakukan mencakup 3 kelas yaitu kelas 1 (0-2 orang bertepuk tangan), kelas 2 (3-7 orang bertepuk tangan), dan kelas 3 (lebih dari 7 orang bertepuk tangan). Tabel 2 menunjukkan rata-rata jumlah sampel yang memiliki nilai amplitudo di atas *threshold* tertentu, yaitu 0.1, 0.2, 0.3, 0.4, dan 0.5 dB untuk masing-masing kelas.

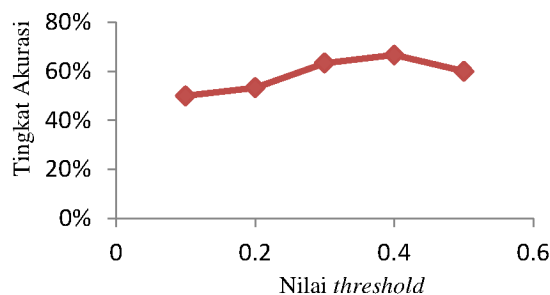
Pengujian langsung dilakukan sebanyak 10 kali untuk masing-masing kelas. Untuk setiap data uji, dilakukan perhitungan jumlah sampel yang memiliki nilai amplitudo di atas nilai-nilai *threshold* yang telah ditentukan. Selanjutnya, jumlah sampel tersebut dibandingkan

dengan model kelas seperti yang ditunjukkan pada Tabel 2. Data uji diklasifikasikan berdasarkan selisih minimum antara jumlah sampel data uji dengan jumlah sampel pada model masing-masing kelas.

Gambar 9 menunjukkan perbandingan hasil akurasi berdasarkan nilai *threshold* yang dipilih. Nilai *threshold* 0.4 dB menghasilkan akurasi maksimum, sedangkan nilai *threshold* 0.1 dB menghasilkan akurasi minimum yaitu 50%. Hal ini menunjukkan bahwa nilai *threshold* 0.1 tidak dapat mewakili suara tepuk tangan secara jelas. Nilai *threshold* 0.5 dB juga tidak menunjukkan akurasi sebaik nilai *threshold* 0.4. Hal ini bisa terjadi disebabkan, nilai 0.5 terlalu tinggi untuk dijadikan batas sehingga ciri suara tidak dapat dikenal dengan baik. Tabel 3 menunjukkan bahwa akurasi sistem berbasis amplitudo dengan nilai *threshold* 0.4 dB adalah 66.7%.

Tabel 2 Pemodelan jumlah sampel tiap kelas berdasarkan nilai *threshold* amplitudo

Jumlah Sampel	Amplitudo (dB)				
	≥ 0.1	≥ 0.2	≥ 0.3	≥ 0.4	≥ 0.5
Kelas 1	1338	473	202	77	36
Kelas 2	2176	887	471	312	246
Kelas 3	3440	1334	786	492	321



Gambar 9 Perbandingan hasil akurasi berdasarkan nilai *threshold*

Tabel 3 Matriks konfusi hasil pengujian langsung data tepuk tangan acak dengan *threshold* 0.4

Prediksi \ Aktual	Kelas 1	Kelas 2	Kelas 3
Kelas 1	7	3	0
Kelas 2	3	6	1
Kelas 3	0	3	7

Keterangan: Akurasi = $\frac{20}{30} \times 100\% = 66.7\%$

SIMPULAN DAN SARAN

Implementasi metode berbasis frekuensi dengan MFCC dan *Codebook* menunjukkan bahwa nilai-nilai parameter yang dipilih sangat mempengaruhi akurasi sistem. Semakin besar jumlah kluster pada tahap pemodelan kelas, maka nilai akurasi sistem akan semakin baik.

Hasil evaluasi sistem menunjukkan bahwa metode *threshold* yang berbasis amplitudo menghasilkan akurasi 66.7%. Hasil ini lebih baik dibandingkan dengan metode MFCC berbasis frekuensi yang hanya menghasilkan akurasi 50%. Hal ini menunjukkan bahwa metode berbasis frekuensi tidak cocok diterapkan untuk aplikasi perhitungan banyaknya orang yang bertepuk tangan, namun lebih cocok untuk mengidentifikasi suatu jenis suara tertentu, misalnya untuk membedakan suara manusia. Untuk aplikasi perhitungan suara tepuk tangan, metode yang lebih cocok diterapkan ialah metode berbasis amplitudo.

DAFTAR PUSTAKA

- Buono A. 2009. Representasi nilai HOS dan model MFCC sebagai ekstraksi ciri pada sistem identifikasi pembicara di lingkungan ber-*noise* menggunakan HMM. [Disertasi]. Depok (ID): Universitas Indonesia.
- Buono A, Kusumoputro B. 2007. Pengembangan model HMM berbasis maksimum lokal menggunakan jarak Euclid untuk sistem identifikasi pembicara. Di dalam: *Prosiding pada workshop NACSIIT*; 2007 Jan 29-30; Depok, Indonesia. hlm 49-54.
- Do MN. 1994. *Digital Signal Processing Mini-Project: An Automatic recognition System. Audio Visual Communication Laboratory*, Swiss Federal Institute of Technology, Switzerland.
- Ganchev T. 2005. Speaker recognition. [Disertasi]. Patras (GR): Department of Computer and Electrical Engineering, University of Patras.
- Li TF, Chang SC. 2003. *Speech Recognition of Mandarin Syllables Using both LPCC and MFCC*. [Paper]. Institute of Management Ming Dao University, Department of Information Management The Overseas Chinese Institute of Technology.
- Nilsson M, Ejnarsson M. 2002. Speech recognition using Hidden Markov Model: performance evaluation in noisy environment. [Tesis]. Blekinge (SE): Blekinge Institute of Technology.
- Taufani MF. 2011. Perbandingan Pemodelan *Wavelet* dan *MFCC* sebagai Ekstraksi Ciri pada Pengenalan Fonem dengan Teknik Jaringan Syaraf Tiruan sebagai *Classifier*. [Skripsi]. Bogor (ID): Institut Pertanian Bogor.
- Wisnudisastra E, Buono A. 2010. Pengenalan *chord* pada alat musik gitar menggunakan *codeBook* dengan teknik ekstraksi ciri MFCC. *Jurnal Ilmiah Ilmu Komputer*. 14(1):16-21.